University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Farokhi, F;Shames, I

Title:
Compressive Sensing in Fault Detection

Date:
2018-08-09

Citation:
Farokhi, F. & Shames, I. (2018). Compressive Sensing in Fault Detection. Proceedings of the American Control Conference, 2018-June, pp.159-164. IEEE. https://doi.org/10.23919/ACC.2018.8431017.

Persistent Link:
https://hdl.handle.net/11343/251420

Compressive Sensing in Fault Detection

Farhad Farokhi and Iman Shames

Abstract— **Randomly generated tests are used to identify faulty sensors in large-scale discrete-time linear time-invariant dynamical systems with high probability. It is proved that the number of the required tests for successfully identifying the location of the faulty sensors (with high probability) scales logarithmically with the number of the sensors and quadratically with the maximum number of faulty sensors. It is also proved that the problem of decoding the identity of the faulty sensors based on the random tests can be cast as a linear programming problem and therefore can be solved reliably and efficiently even for large-scale systems. A numerical example based on automated irrigation networks is utilized to demonstrate the results.**

Index Terms— **Fault detection; Linear programming; Randomized tests; Compressive sensing.**

I. Introduction

Advances in network engineering and communication systems have enabled the use of wireless communication networks in the control of physical systems. These networks transmit the sensor measurements from the sensors to the controller and the control actions from the controller to the actuators. Although flexible and easy to maintain, adoption of wireless communication systems introduces new avenues for faults into the system. For instance, the communication link between a given sensor and the controller might break down or an adversary might perform sophisticated attacks on a subset of the communication channels to jeopardise the performance of the closed-loop system. Furthermore, advances in control engineering has enabled us to design and deploy larger cyber-physical systems consisting of many individual elements working in synchrony. It is evident that, at any moment, a fault might exhibit itself in one of the elements. Therefore, a systematic approach is required for monitoring the communication channels, the sensors, and the actuators to identify faulty or mischievous elements.

A common method for fault detection is to generate residuals using state estimators, unknown input observers, or system identification methods under various faults scenarios. Following that, the most likely faults can be identified using a maximum likelihood test. For instance, if it is desired to find a faulty sensor, a bank of Kalman filters can be designed where each Kalman filter uses all but one of the sensors. Here, it is assumed that there is enough redundancy in the output measurements so that the system remains observable in the absence of one sensor reading. Evidently, the residuals of the Kalman filters containing faulty sensors behaves differently than of the rest. This way, the faulty sensors can be detected. A brute-force approach consists of checking all the sensors separately. However, group testing methods have been also developed to identify faulty sensors [1]. The group testing approach shows that the number of Kalman filters that need to be developed is in fact far lower than the number of the sensors. Therefore, computational resources can be saved.

Group testing was developed as a method for testing individual with rare diseases in a combined manner (by mixing their blood samples) to identify the infected individuals using fewer tests (in comparison to testing them individually) [2]. Since then, group testing has found its way to many applications, such as communication engineering [3], data forensic [4], DNA analysis [5], and fault detection [6]. with the exception of [1], group testing has not been used in fault detection in dynamical systems. Furthermore, in [1], the construction of the set of tests (for group testing) and the eventual maximum likelihood test for decoding the test results are combinatorial and are thus computationally intensive.

In this paper, it is proved that randomly generated tests can obtain the same result with a high probability. In fact, the probability of failure can be reduced by increasing the number of tests. It is proved that the number of the required tests (for successfully detecting the faulty sensors with a high probability) scales logarithmically with the number of the sensors and quadratically with the maximum number of faulty sensors. With the aid of the results of [7], in this paper, it is also proved that the problem of decoding the faulty sensors based on the random tests can be cast as a linear programming problem and therefore can be solved reliably and efficiently even for large scale systems.

This problem is in close relationship to compressive sensing [8], [9], where few measurements are used to reconstruct a large, yet sparse (in some domain), vector of unknowns. Compressive sensing has been more recently applied to group testing [10], [11]. However, these results have not been utilized within the context of fault detection and cybersecurity, which is the topic of this paper. Compressive sensing has been also utilized within control community [12]–[14]. However, the focus of those results has been on the observability and the state estimation of sparsely-representable dynamical systems.

The rest of the paper is organized as follows. Section II explains the problem formulation. The results are presented in Section III. Finally, numerical results are presented in

F. Farokhi and I. Shames are with the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia. e-mail: {ffarokhi,ishames}@unimelb.edu.au

The work of F. Farokhi was supported by the McKenzie Fellowship from the University of Melbourne and the veski Victoria Fellowship from the Victorian State Government.

The work of I. Shames was supported by Discover Project DP170104099.

Section IV and the paper is concluded in Section V.

II. Problem Formulation

Consider the discrete-time linear time-invariant[1] dynamical system

$$x[k+1] = Fx[k] + Gu[u] + w[k],$$
$$y[k] = Hx[k] + v[k] + f[k],$$

where $x[k] \in \mathbb{R}^n$ is the state of the system, $u[k] \in \mathbb{R}^m$ is the control input, $w[k]$ is the process noise (modelling the effects of stochastic exogenous inputs and model uncertainties), $y[k] \in \mathbb{R}^p$ is the output measurement, $v[k] \in \mathbb{R}^p$ is the measurement noise (modelling the effect of sensing uncertainties and imperfections), and $f[k] \in \mathbb{R}^p$ models the sensing faults or malicious behaviours. It is assumed that $w[k]$ and $v[k]$ are independent i.i.d.[2] zero mean Gaussian variables with covariance Q and R, respectively.

A Kalman filter with a subset of measurements denoted by $y_{\mathcal{K}}[k] := (y_i[k])_{i\in\mathcal{K}}$ with $\mathcal{K} \subseteq \{1,\dots,p\}$ can be used to estimate the states of the system. The prediction step for the Kalman filter for this combination of outputs is given by

$$\hat{x}_{\mathcal{K}}[k|k-1] = F\hat{x}_{\mathcal{K}}[k-1|k-1] + Gu[k-1],$$
$$P_{\mathcal{K}}[k|k-1] = FP_{\mathcal{K}}[k-1|k-1]F^\top + Q,$$

where the estimates $\hat{x}_{\mathcal{K}}[k|k-1]$ and $\hat{x}_{\mathcal{K}}[k-1|k-1]$, respectively, denote $\mathbb{E}\{x[k] \,|\, y_{\mathcal{K}}[0],\dots,y_{\mathcal{K}}[k-1]\}$ and $\mathbb{E}\{x[k-1] \,|\, y_{\mathcal{K}}[0],\dots,y_{\mathcal{K}}[k-1]\}$, and $P_{\mathcal{K}}[k-1|k-1]$ and $P_{\mathcal{K}}[k|k-1]$, respectively, denote the variance of $\mathbb{E}\{x[k-1] \,|\, y_{\mathcal{K}}[0],\dots,y_{\mathcal{K}}[k-1]\}$ and $\mathbb{E}\{x[k] \,|\, y_{\mathcal{K}}[0],\dots,y_{\mathcal{K}}[k-1]\}$. The update step of the Kalman filter is also given by

$$\hat{x}_{\mathcal{K}}[k|k] = \hat{x}_{\mathcal{K}}[k|k-1] + L_{\mathcal{K}}[k](y_{\mathcal{K}}[k] - H_{\mathcal{K}}\hat{x}_{\mathcal{K}}[k|k-1]),$$
$$L_{\mathcal{K}}[k] = P_{\mathcal{K}}[k|k-1]H_{\mathcal{K}}^\top(H_{\mathcal{K}}P_{\mathcal{K}}[k|k-1]H_{\mathcal{K}}^\top + R_{\mathcal{K}})^{-1},$$
$$P_{\mathcal{K}}[k|k] = (I - L_{\mathcal{K}}[k]H_{\mathcal{K}})P_{\mathcal{K}}[k|k-1],$$

where $H_{\mathcal{K}}$ is a matrix containing subset of the rows of H corresponding to the sensor measurements in the set $\mathcal{K}$ and $R_{\mathcal{K}}$ is the co-variance of $v_{\mathcal{K}}[k]$, i.e., the part of the measurement noise $v[k]$ contained in the contributing to the selected outputs $y_{\mathcal{K}}$. Now, if the pair $(F, H_{\mathcal{K}})$ is observable and $f_{\mathcal{K}}[0] = 0$ (i.e., the outputs in $\mathcal{K}$ are fault free), the residual of this Kalman filter $y_{\mathcal{K}}[k] - H_{\mathcal{K}}\hat{x}_{\mathcal{K}}[k|k]$ is an i.i.d. Gaussian random variable with zero mean and covariance $H_{\mathcal{K}}P_{\mathcal{K}}[k|k]H_{\mathcal{K}}^\top + R$ [15]. Therefore, the χ^2-test or the Neyman-Pearson test can be utilized to figure out if the output measurements in $\mathcal{K}$ contain any faulty sensors to not.

Define $z_i \in \{0,1\}$ to be a boolean variable such that $z_i = 0$ if sensor i is not faulty (i.e., $f_i[k] = 0$ for all k) and $z_i = 1$ otherwise. The Kalman filter approach above can be used to compute

$$\xi_{\mathcal{K}} = \left(\bigvee_{j\in\mathcal{K}} z_j\right) \vee \zeta_{\mathcal{K}}, \tag{1}$$

where $\zeta_{\mathcal{K}}$ is a Bernoulli random variable with mean q (i.e., the probability of the event $\zeta_{\mathcal{K}} = 1$ is equal to q) capturing the false detection. Note that, without loss of generality, the tests can be fine tuned so that they have the same probability of false detections.

Problem 1: Find a family of tests $\mathcal{K}_t$, $t = 1,\dots,T$, such that the set of faulty sensors can be reconstructed from the noisy boolean measurements of the form (1) with a high probability.

Now that problem formulation has been formalized, we can present the results of the paper. This is the subject of the next section.

III. Main Results

For a set of tests $\mathcal{K}_t$, $t = 1,\dots,T$, define $A \in \{0,1\}^{T\times p}$ such that a_{ij}, i.e., the entry in i-th row and j-th column of A, is equal to one if and only if $j \in \mathcal{K}_i$. Hence, the boolean expression in (1) for a given test set $\mathcal{K}_i$ can be rewritten as

$$\xi_{\mathcal{K}_i} = \left(\bigvee_{j=1}^{p} (a_{ij} \wedge z_j)\right) \vee \zeta_{\mathcal{K}_i}.$$

Note that there is an equivalent between each row of the matrix A and each test $\mathcal{K}_i$. Therefore, these entities can be used interchangeably. Following [7], recovering the exact vector of faults $z := (z_i)_{i=1}^p$ can be achieved through solving the following non-linear integer program:

$$(\hat{z}, \hat{\zeta}) \in \arg\min_{\bar{z},\bar{\zeta}} \|\bar{z}\|_0 + \alpha\|\bar{\zeta}\|_0, \tag{2a}$$
$$\text{s.t.} \quad \bar{z} \in \{0,1\}^p, \tag{2b}$$
$$\bar{\zeta} \in \{0,1\}^T, \tag{2c}$$
$$\xi_{\mathcal{K}_i} = \left(\bigvee_{j=1}^{p} (a_{ij} \wedge \bar{z}_j)\right) \vee \bar{\zeta}_i, \quad \forall i \in \{1,\dots,T\}, \tag{2d}$$
$$\|\bar{z}\|_0 \le d, \tag{2e}$$

where $d \in \mathbb{N}$ denotes the maximum number of faults that can be present and $\alpha > 0$ is a constant balancing between the number of faults $\|\bar{z}\|_0$ and the number of the miss identifications $\|\bar{\zeta}\|_0$. For the noiseless case, that is if $\zeta_i = 0$, it can be shown that the optimization problem (2) can be transformed into:

$$\hat{z} \in \arg\min_{\bar{z}} \|\bar{z}\|_0, \tag{3a}$$
$$\text{s.t.} \quad \bar{z} \in \{0,1\}^p \tag{3b}$$
$$\xi_{\mathcal{K}_i} = \bigvee_{j=1}^{p} (a_{ij} \wedge \bar{z}_j), \forall i \in \{1,\dots,T\}, \tag{3c}$$
$$\|\bar{z}\|_0 \le d, \tag{3d}$$

The optimization problem (3) can be utilized in this case to find the true vector of faults $z := (z_i)_{i=1}^p$ if $T = (e\log 2)d^2\log(p)$ [7], [16]. In this case, in fact a combinatorial method must be utilized to carefully construct a matrix A (or the tests $\mathcal{K}_t$, $t = 1,\dots,T$ equivalently) to ensure that the faults are always identifiable. Furthermore, a necessary

[1] The assumption regrading time invariant nature of the dynamical systems is for the sake of simplicity of the presentation of the results and can be easily removed.

[2] i.i.d. stands for independently and identically distributed.

condition on the number of the required tests for identification of the faults is $T \geq d^2 \log(p)/\log(e(d+1)/2)$ [7], [16]. Although powerful at reducing the number of required tests (in comparison to a brute-force approach), this method however is impractical for large-scale system due to two reasons. First, the computational complexity of designing the tests is high. Secondly, the optimization problem (3) is difficult to solve because of (*i*) the integer constraints in (3b), (*ii*) the non-linear nature of the equality constraint in (3c) and the inequality constraint in (3d), and (*iii*) the non-differentiable non-convex nature of the cost function. Therefore, this method can only be utilized for small systems with few faults (which is not in line with the motivation behind the use of group testing).

To alleviate some of these issues, the optimization problem in (3) can be relaxed (or convexified) to get

$$\hat{z} \in \arg\min_{\bar{z}} \|\bar{z}\|_1, \tag{4a}$$
$$\text{s.t.} \quad \bar{z} \in [0,1]^p, \tag{4b}$$
$$A_i \bar{z} = 0, \forall i \in \{1,\dots,T\} : \xi_{\mathcal{K}_i} = 0 \tag{4c}$$
$$A_i \bar{z} \geq 1, \forall i \in \{1,\dots,T\} : \xi_{\mathcal{K}_i} = 1 \tag{4d}$$
$$\|\bar{z}\|_1 \leq d. \tag{4e}$$

Before moving any further, a few definitions must be presented. For any two boolean vectors $a, b \in \{0,1\}^n$, the union of a and b is given by a boolean vector $c \in \{0,1\}^n$ in which $c_j = a_j \vee b_j$ for all $1 \leq j \leq n$. Further, a is said to be contained by b if $\{i \,|\, a_i = 1\} \subseteq \{i \,|\, b_i = 1\}$.

Definition 1 (k-disjunct): A boolean matrix A is called k-disjunct if the union of any k columns of A does not contain any other column from A.

This condition for boolean vectors is similar at heart to the linear independence. A boolean matrix is k-disjunct if no group of any k columns from it can reconstruct any other of its columns.

Proposition 1: The solution of (4) is a solution of (3) if A is d-disjunct.

Proof: The proof is a direct application of [7]. ∎

Proposition 1 shows that the the relaxation in (4) takes away the complexity of solving (3) while not introducing any conservatism in terms of identifying faulty sensors if the measurement matrix (or the corresponding family of tests) is d-disjunct. However, the procedure for forming matrix A to be d-disjunct still remains complex.

In what follows a random procedure for constructing the family of the tests so that the measurement matrix meets the condition of Proposition 1 is considered. Let a randomly generated matrix A be generated such that a_{ij} is a Bernoulli random variable with mean γ (i.e., the probability of the event $a_{ij} = 1$ is equal to γ). For these random matrices, the following result can be proved.

Proposition 2: $\mathbb{P}\{A \text{ is } d\text{-disjunct}\} \geq 1 - 1/p^\delta$ for all $T \geq (16 + 8\delta) d^2 \ln(p)$.

Proof: Let A^j denote the j-th column of A. We claim that if $\min_j |\{i \,|\, A_i^j = 1\}| \geq \alpha$ and $\max_{j_1 \neq j_2} |\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \leq \beta$, then A is $\lfloor(\alpha-1)/\beta\rfloor$-disjunct.

To prove this claim, let us select an arbitrary set of $\lfloor(\alpha-1)/\beta\rfloor$ columns A^j from matrix A. Now, any other column of matrix A, denoted by A^ℓ, can share at most β non-zero elements with any column from our selection. Noting that there are $\lfloor(\alpha-1)/\beta\rfloor$ columns in the selection, A^ℓ can at most share $\beta\lfloor(\alpha-1)/\beta\rfloor \leq \alpha - 1$ non-zero elements with the selection. Noting that A^ℓ has at least α non-zero entries, there exists at least one non-zero element in A^ℓ that cannot be matched by any of the columns from the selection. Therefore, A^ℓ cannot be contained by the selection of the columns. Hence, the matrix A is $\lfloor(\alpha-1)/\beta\rfloor$-disjunct. In the remainder of this proof, we show that these conditions are met with a high probability for the random matrix A.

By Chernoff bound [17, Ch. 4], it can be deduced that $\mathbb{P}\{|\{i \,|\, A_i^j = 1\}| \leq \gamma T/2\} \leq \exp(-\gamma T/8)$ because $\mathbb{E}\{|\{i \,|\, A_i^j = 1\}|\} = \gamma T$. This shows that

$$\begin{aligned}
&\mathbb{P}\{\exists j : |\{i \,|\, A_i^j = 1\}| \leq \gamma T/2\} \\
&\quad \leq \sum_{j=1}^{p} \mathbb{P}\{|\{i \,|\, A_i^j = 1\}| \leq \gamma T/2\} \\
&\quad = p\mathbb{P}\{|\{i \,|\, A_i^j = 1\}| \leq \gamma T/2\} \\
&\quad \leq p \exp(-\gamma T/8).
\end{aligned}$$

Similarly, with the aid of Chernoff bound [17, Ch. 4], it can be deduced that $\mathbb{P}\{|\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \geq 2\gamma^2 T\} \leq \exp(-\gamma^2 T/3)$ because $\mathbb{E}\{|\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}|\} = \gamma^2 T$. Therefore, it can be proved that

$$\begin{aligned}
&\mathbb{P}\{\exists j_1, j_2 : |\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \geq 2\gamma^2 T\} \\
&\leq \sum_{j_1 \neq j_2} \mathbb{P}\{\exists j_1, j_2 : |\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \geq 2\gamma^2 T\} \\
&\leq p(p-1)\exp(-\gamma^2 T/3) \\
&\leq p^2 \exp(-\gamma^2 T/3).
\end{aligned}$$

Finally, with the aid of these two inequalities, it can be shown that

$$\begin{aligned}
&\mathbb{P}\Big\{ \max_{j_1 \neq j_2} |\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \leq 2\gamma^2 T \\
&\qquad \text{and } \min_j |\{i \,|\, A_i^j = 1\}| \geq \gamma T/2 \Big\} \\
&= 1 - \mathbb{P}\big\{\exists j_1, j_2 : |\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \geq 2\gamma^2 T \\
&\qquad\qquad \text{or } \exists j : |\{i \,|\, A_i^j = 1\}| \geq \gamma T/2 \big\} \\
&\geq 1 - (\mathbb{P}\{\exists j_1, j_2 : |\{i \,|\, A_i^{j_1} = 1\} \cap \{i \,|\, A_i^{j_2} = 1\}| \geq 2\gamma^2 T\} \\
&\qquad + \mathbb{P}\{\exists j : |\{i \,|\, A_i^j = 1\}| \geq \gamma T/2\}) \\
&\geq 1 - p^2 \exp(-\gamma^2 T/3) - p \exp(-\gamma T/8).
\end{aligned}$$

Now, notice that

$$\left\lfloor \frac{\gamma T/2 - 1}{2\gamma^2 T} \right\rfloor \geq \frac{\gamma T/2 - 1}{2\gamma^2 T} - 1.$$

Hence, $\lfloor(\gamma T/2 - 1)/(2\gamma^2 T)\rfloor \geq d$ (recall that A must be d-disjunct for a successful recovery), we must guarantee that

$$\frac{\gamma T/2 - 1}{2\gamma^2 T} - 1 \geq d.$$

For $\gamma = 1/d$, this condition is equivalent to $T \geq 2d^2/(5d+4)$. A sufficient condition for this inequality is to select $T \geq 2d/5$ because $2d/5 \geq 2d^2/(5d+4)$ for all $d \geq 0$. Thus, $\mathbb{P}\{A \text{ is } d\text{-disjunct}\} \geq 1 - p^2\exp(-\gamma^2 T/3) - p\exp(-\gamma T/8)$ if $T \geq 2d/5$.

Now, note that if $T = cd^2\ln(p)$ and if $\gamma = 1/d$, it can be seen that $\exp(-\gamma T/8) = \exp(-cd\ln(p)/8) = p^{-cd/8}$. Thus, $p\exp(-\gamma T/8) \leq 1/(2p^\delta)$ if $c \geq 8(\log_p(2) + \delta + 1)/d$. Similarly, if $T = cd^2\ln(p)$ and $\gamma = 1/d$, we get $\exp(-\gamma^2 T/3) = \exp(-c\ln(p)/3) = p^{-c/3}$. Therefore, $p^2\exp(-\gamma^2 T/3) \leq 1/(2p^\delta)$ if $c \geq 3(2+\delta+\log_p(2))$. Combining all these inequalities shows that $\mathbb{P}\{A \text{ is } d\text{-disjunct}\} \geq 1 - 1/p^\delta$ if $c \geq 8\delta + 8 + 8\log_p(2)$. Finally note that $\log_p(2) = \ln(2)/\ln(p) \leq \ln(2)/\ln(2) = 1$ for all $p \geq 2$. Therefore, $\mathbb{P}\{A \text{ is } d\text{-disjunct}\} \geq 1 - 1/p^\delta$ if $c \geq 16 + 8\delta$. This concludes the proof. ■

Theorem 1: The solution of (4) is a solution of (3) with at least probability $1/p^\delta$ if $T \geq 24d^2\ln(p)$.

Proof: The proof follows from the application of the results of Propositions 1 and 2. ■

It should be noted as opposed the deterministic algorithms for generating matrix A, in the stochastic case, there is a non-zero probability for not being able to recover the faulty outputs correctly. However, this probability can be pushed arbitrarily close to zero.

We need to also study the observability of the system using the random test sets. To do so, we need to define a sense of redundancy among the sensors.

Definition 2 (k-observable): A pair of matrices (F, H) is k-observable if $(F, H_\mathcal{K})$ is observable for any $\mathcal{K}$ with cardinality k.

Again, we would like to mention that there is an equivalence between A and the family of tests $(\mathcal{K}_i)_{i=1}^T$. Therefore, these entities can be used interchangeably.

Proposition 3: Let (F, H) be k-observable. Then $\mathbb{P}\{(A, H_\mathcal{K}) \text{ is observable}\} \geq 1 - 1/p^\delta$ for a Bernoulli test $\mathcal{K}$ (corresponding to the rows of the matrix A) if $k \leq p/d - 2\delta\ln(p)$.

Proof: For a given test $\mathcal{K}$, a sufficient condition for $(F, H_\mathcal{K})$ to be observable is to have $|\mathcal{K}| \geq k$. The Chernoff bound [17, Ch. 4] shows that

$$\mathbb{P}\{|\mathcal{K}| \geq k\} \geq 1 - \exp\left(-\frac{(p-kd)^2 p}{2dp}\right), \forall k \leq \frac{p}{d}.$$

To ensure $\mathbb{P}\{|\mathcal{K}| \geq k\} \geq 1 - 1/p^\delta$, we must ensure that

$$\exp\left(-\frac{(p-kd)^2 p}{2dp}\right) \leq \frac{1}{p^\delta}.$$

This condition is satisfied if $k \leq p/d - 2\delta\ln(p)$. This concludes the proof. ■

Proposition 3 shows that to ensure observability with the randomly selected outputs, the systems needs to have a high level of redundancy if $d > 0$. For instance, for $d = 2$, the system must be $(p/2 - 2\delta\ln(p))$-observable, i.e., more than half of the sensor measurements must be redundant for observability. However, the condition implies that the system must be $(p - 2\delta\ln(p))$-observable for $d = 1$, which is not extremely conservative. Reducing the conservatism of this condition remains an important avenue for future research.

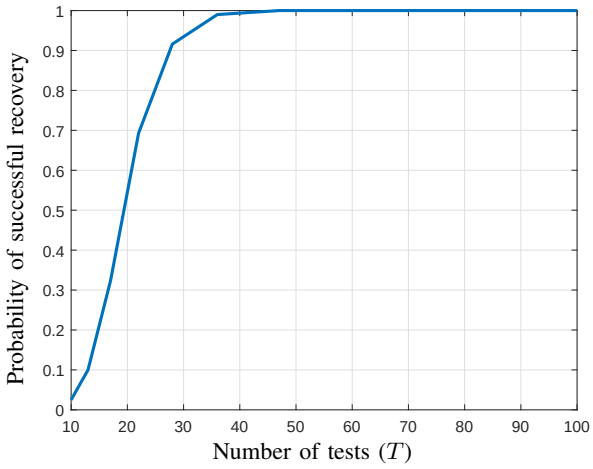


Fig. 1. Probability of successful recovery as a function of the number of the tests T for the noiseless case with $p = 100$.

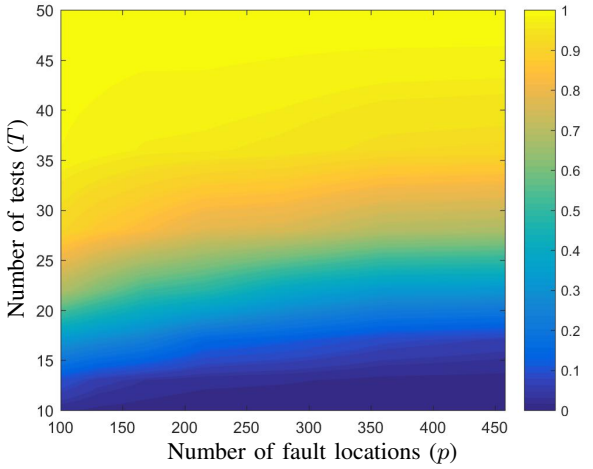


Fig. 2. Probability of successful recovery (captured by the colours) as a function of the number of the tests T for the noiseless case with $p = 100$.

For noisy cases, i.e., if $\zeta_{\mathcal{K}_i}$ are non-zero with a non-zero probability, the optimization problem (4) is no longer useful. In this case, the following alternative must be solved:

$$(\hat{z}, \hat{\zeta}) \in \arg\min_{\bar{z}, \bar{\zeta}} \|\bar{z}\|_1 + \alpha\|\bar{\zeta}\|_1, \tag{5a}$$

$$\text{s.t.} \quad \bar{z} \in [0,1]^p, \tag{5b}$$

$$\bar{\zeta} \in \mathbb{R}^p_{\geq 0}, \tag{5c}$$

$$A_i\bar{z} = \bar{\zeta}_i, \forall i : \xi_{\mathcal{K}_i} = 0 \tag{5d}$$

$$A_i\bar{z} + \bar{\zeta}_i \geq 1, \bar{\zeta}_i \leq 1, \forall i : \xi_{\mathcal{K}_i} = 1 \tag{5e}$$

$$\|\bar{z}\|_1 \leq d. \tag{5f}$$

In the next section, we demonstrate the applicability of these results over a large scale dynamical system.

IV. Numerical Example

The applicability of the developed algorithms is demonstrated on a water channel with N pools. The model and the parameters of the pools are borrowed from [18], [19]. Each pool can be modelled by

$$y_i(s) = \frac{c_{\text{in},i}}{s} e^{-t_{\text{d},i}s} h_i(s) - \frac{c_{\text{out},i}}{s} h_{i+1}(s) + \frac{c_{\text{out},i}}{s} \zeta_i(s),$$

where $c_{\text{in},i}$ and $c_{\text{out},i}$ are the discharge rates of the gates between neighbouring pools, $t_{\text{d},i}$ is the delay associated with the transport of water along each pool, $\zeta_i(s)$ is the overall off-take flow load on pool i, $h_i(s)$ is the flow of water from pool $i-1$ to pool i, and $y_i(s)$ denotes the water level in pool i. The transfer function of the delay is replaced with its first-order Padé approximation. This is due to that the closed-loop dynamics of the channel with local PI[3] controllers is insensitive to the open-loop modelling error associated with the Padé approximation of the delay (because the loop-gain at frequencies where the approximation error is significant is small) [20]. Each pool is assumed to be locally controlled by

$$h_i(s) = \frac{\kappa_i(\eta_i s + 1)}{s(\rho_i s + 1)}(u_i(s) - y_i(s)),$$

where κ_i, η_i, and ρ_i are appropriately selected control parameters, and $u_i(s)$ is the reference signal of pool i. The aggregate model of the system is discretized with the sampling time of $30\,\text{sec}$ (or $0.5\,\text{min}$) with a zero-order hold on the reference signals. Following that, the model of the whole system is given by

$$x[k+1] = Ax[k] + Bu[k] + w[k],$$

where $w[k]$ is a process noise modelling the effect of local disturbances, such as off-take uncertainties, and physical phenomenons, such as wind. The variance of the process noise is set as $Q = 0.01I$. The outputs of the system $y[k]$ are the measurements of water level at all the pool and the water flow among all the pools. The variance of the measurement is equal to $R = 0.0025I$. We consider the case where the sensors of the pool can simultaneously drop, e.g., due to communication issues. Thus $p = N$ (as there are N pools). Furthermore, it is assumed that the maximum number of sensor faults is $d = 2$. Finally, to ensure observability of the system, it is assumed that each test involves the measurements of the last pool. This can be motivated by that the last pool is monitored separately to see if its sensors are faulty or not while the rest of the pools being remote locations cannot be accessed for regular service.

A. Noiseless Measurements

We start by demonstrating the noiseless case divorced from the underlying dynamical system. Consider the case where $p = 100$ and noiseless measurements of the form (1) are available. Further, it is assumed that two simultaneous faults are present. Figure 1 illustrates the probability of successfully detecting the faulty measurements (empirically estimated using 1000 simulations) versus number of tests. Evidently the probability of successful recovery improves with increasing the number of the tests. There is a threshold after the linear programming algorithm almost surely recover the faulty entries almost surely. Figure 2 shows the probability of success versus both the number of the tests T and the number of fault locations p. As expected the number of the required tests to be able to extract the location of the faults reliably grows slowly with the number of the fault locations p.

[3]PI stands for proportional-integral.

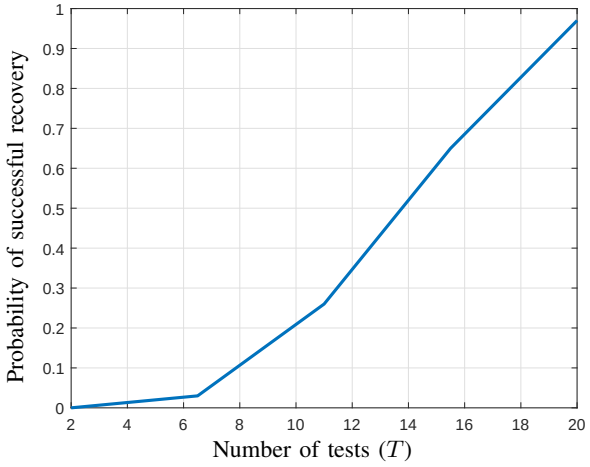


Fig. 3. Probability of successful recovery as a function of the number of the tests T for the noisy case with $p = 20$ and χ^2-test over a window of 10 samples.

B. Noisy Measurements

Consider the case there are $N = 20$ pools in the irrigation network. We use the Kalman filters in Section II to investigate if the sensors in each pool are faulty or not. We use the χ^2-test over a window of 10 samples to decide if there is a fault present or not, i.e., it is deemed that a fault is present if, for some k,

$$\frac{1}{20}\sum_{t=0}^{9}(y_{\mathcal{K}}[k-t] - H_{\mathcal{K}}\hat{x}_{\mathcal{K}}[k-t|k-t])^\top \times (H_{\mathcal{K}}P_{\mathcal{K}}[k-t|k-t]H_{\mathcal{K}}^\top + R)^{-1} \times (y_{\mathcal{K}}[k-t] - H_{\mathcal{K}}\hat{x}_{\mathcal{K}}[k-t|k-t]) \geq 2.$$

In what follows, the faulty outputs are determined using the linear programming problem in (5) with $\alpha = 10$.

Figure 3 illustrates probability of successful recovery (empirically estimated using 1000 simulations) as a function of the number of the tests T for the noisy case with $p = 20$. As expected, the probability of successful recovery is an increasing function of T. However, the probability of successful recover does not reach one even for $T = 20$. This is because no matter what there is always false detection due to stochastic nature of the problem and the use of the Kalman filters for fault detection.

V. Conclusions and Future Work

In this paper, randomly generated tests were used to identify faulty sensors with a high probability. This method requires fewer tests required tests for successfully detecting the faulty sensors with a high probability in comparison to the brute-force method of checking all the sensors separately. In fact, the number of the tests scales logarithmically with the number of the sensors and quadratically with the maximum number of faulty sensors. The identification of the faulty sensors can also be done by solving a linear program.

References

[1] C. Lo, M. Liu, J. P. Lynch, and A. C. Gilbert, "Efficient sensor fault detection using combinatorial group testing," in *Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 199–206, 2013.

[2] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, 1943.

[3] T. Berger, N. Mehravari, D. Towsley, and J. Wolf, "Random multiple-access communication and group testing," *IEEE Transactions on Communications*, vol. 32, no. 7, pp. 769–779, 1984.

[4] A. De Bonis and G. Di Crescenzo, "Combinatorial group testing for corruption localizing hashing," in *17th Annual International Conference, COCOON 2011, Dallas, TX, USA, August 14-16, 2011. Proceedings* (B. Fu and D.-Z. Du, eds.), Lecture Notes in Computer Science, pp. 579–591, 2011.

[5] H. Q. Ngo and D.-Z. Du, "A survey on combinatorial group testing algorithms with applications to DNA library screening," *Discrete Mathematical Problems with Medical Applications*, vol. 55, pp. 171–182, 2000.

[6] N. J. Harvey, M. Patrascu, Y. Wen, S. Yekhanin, and V. W. Chan, "Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pp. 697–705, IEEE, 2007.

[7] D. Malioutov and M. Malyutov, "Boolean compressed sensing: LP relaxation for group testing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3305–3308, 2012.

[8] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE signal processing magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[9] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[10] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1880–1901, 2012.

[11] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3019–3035, 2014.

[12] M. B. Wakin, B. M. Sanandaji, and T. L. Vincent, "On the observability of linear systems from random, compressive measurements," in *Proceedings of the 49th IEEE Conference on Decision and Control*, pp. 4447–4454, 2010.

[13] H. Ohlsson, M. Verhaegen, and S. S. Sastry, "Nonlinear compressive particle filtering," in *Proceedings of the 52nd Annual Conference on Decision and Control*, pp. 7054–7059, 2013.

[14] W. Dai and S. Yüksel, "Observability of a linear system under sparsity constraints," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2372–2376, 2013.

[15] B. D. Anderson and J. B. Moore, "Optimal filtering," *Prentice Hall*, 1979.

[16] A. Dyachkov and V. Rykov, "A survey of superimposed code theory," *Problems of Control and Information Theory*, vol. 12, no. 4, pp. 1–13, 1983.

[17] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[18] J. Alende, Y. Li, and M. Cantoni, "A $\{0, 1\}$ linear program for fixed-profile load scheduling and demand management in automated irrigation channels," in *Proceedings of the 48th IEEE Conference on Decision and Control held jointly with the 28th Chinese Control Conference*, pp. 597–602, 2009.

[19] F. Farokhi, M. Cantoni, and I. Shames, "Scheduling rigid demands on continuous-time linear shift-invariant systems," in *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, pp. 5358–5363, 2015.

[20] M. Cantoni, E. Weyer, Y. Li, S. K. Ooi, I. Mareels, and M. Ryan, "Control of large-scale irrigation networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 75–91, 2007.